

GAO

Report to the Honorable
William V. Roth and the Honorable
Charles E. Grassley, U.S. Senate

October 1997

TEST AND EVALUATION

Impact of DOD's Office of the Director of Operational Test and Evaluation



19971030 053

DRG QUALITY INFORMATION 6

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

National Security and
International Affairs Division

B-276799

October 24, 1997

The Honorable William V. Roth
The Honorable Charles E. Grassley
United States Senate

The Department of Defense (DOD) has proposed that the practices and policies of the Office of the Director of Operational Test and Evaluation (DOT&E) be modified to reduce the time and cost of developing and fielding new weapon systems. To help focus deliberations on DOD's proposal, you asked us to review DOT&E's operations and organizational structure for overseeing operational testing. Specifically, you asked us to assess (1) DOT&E's efforts and their impact on the quality of operational testing and evaluation¹ in DOD and (2) the strengths and weaknesses of the current organizational framework in DOD for operational testing. As part of our review, we conducted 13 case studies of the testing of individual weapon systems. (Our scope and methodology are described in app. I, and brief descriptions of the 13 weapon systems are provided in app. II.)

Background

In 1983, Congress established DOT&E to coordinate, monitor, and evaluate operational testing of major weapon systems.² As part of the Office of the Secretary of Defense (OSD), DOT&E is separate from the acquisition community that conducts developmental and operational testing and therefore is in a position to provide the Secretary and Congress with an independent view. Congress created DOT&E in response to reports of conflicts of interest in the acquisition community's oversight of operational testing leading to inadequate testing of operational suitability³ and effectiveness⁴ and the fielding of new systems that performed poorly. (DOD's system acquisition process is described in app. III.)

¹The term "operational test and evaluation" means (1) the field test, under realistic conditions, of any item or key component of a weapon system, equipment, or munition for the purpose of determining the effectiveness and suitability of the weapon, equipment, or munition for use in combat by typical military users and (2) the evaluation of the results of the test.

²P.L. 98-94 sec. 1211(a)(1), 97 Stat. 684. DOT&E's legislation is now codified at 10 U.S.C. 139.

³DOD defines "operationally suitable" as the degree to which a system can be placed satisfactorily in field use, with consideration given to such factors as availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, and supportability.

⁴DOD defines "operationally effective" as the overall degree of mission accomplishment of a system when used by representative personnel in the environment planned or expected for operational employment of the system, considering organization, doctrine, tactics, survivability, vulnerability, and threat.

By law, DOT&E serves as the principal adviser on operational test and evaluation in DOD and bears several key responsibilities, including

- monitoring and reviewing all operational test and evaluation in DOD,
- reporting to the Secretary of Defense and congressional committees whether the tests and evaluations of weapon systems were adequate and whether the results confirmed that the system is operationally suitable and effective for combat before a decision is made to proceed to full-rate production, and
- submitting to the Secretary of Defense and congressional decisionmakers an annual report summarizing operational test and evaluation activities during the preceding fiscal year.

In 1993, DOD's advisory panel on streamlining and codifying acquisition laws⁵ concluded that DOT&E was impeding the goals of acquisition reform by (1) promoting unnecessary oversight, (2) requiring excessive reporting detail, (3) inhibiting the services' discretion in testing, and (4) limiting participation of system contractors in operational tests where such involvement is deemed necessary by the services. The following year, DOD proposed legislative changes that would have reduced the scope and authority of DOT&E. In testimony, we opposed these changes because they were directed at perceived rather than documented problems and would undermine a key management control over the acquisition process—Independent oversight of operational test and evaluation.⁶

Although the legislative proposals were not adopted, in 1995 the Secretary of Defense implemented several operational test and evaluation initiatives in the Department to (1) involve operational testers earlier in the acquisition process, (2) use models and simulations effectively, (3) combine tests where possible, and (4) combine tests and training. The goals of these initiatives included saving time and money by identifying and addressing testing issues earlier in the acquisition process; merging or closely coordinating historically distinct phases, such as developmental and operational testing to avoid duplication; and using existing technologies and training exercises to create realistic and affordable test conditions.

⁵Established under section 800 of the National Defense Authorization Act for Fiscal Year 1991 (P.L. 101-510, 1990).

⁶Acquisition Reform: Role of Test and Evaluation in System Acquisition Should Not Be Weakened (GAO/T-NSIAD-94-124, Mar. 22, 1994).

Results in Brief

Our review of 13 case studies indicated that DOT&E oversight of operational testing and evaluation increased the probability that testing would be more realistic and more thorough.⁷ Specifically, DOT&E was influential in advocating increasing the reliability of the observed performance and reducing the risk of unknowns through more thorough testing; conducting more realistic testing; enhancing data collection and analysis; reporting independent findings; and recommending follow-on operational test and evaluation when suitability or effectiveness was not fully demonstrated prior to initiating full-rate production.

The independence of DOT&E—and its resulting authority to report directly to Congress—is the foundation of its effectiveness. That independence, along with its legislative mandate, provides sufficient freedom and authority to exercise effective oversight of the operational testing and evaluation of new systems before a decision is made to begin full-rate production. In the conduct of its oversight, DOT&E (1) executes its approval authority over test and evaluation master plans and operational test plans and (2) provides independent annual and summary reports on the test and evaluation of individual weapon systems to the Secretary of Defense and Congress.

DOT&E can reduce the risk that systems are not adequately tested prior to the full-rate production decision. But DOT&E cannot ensure that (1) only systems whose operational effectiveness and suitability have been demonstrated through operational testing will proceed to the full-rate production decision or (2) new fielded systems will accomplish their missions as intended or that the fielded systems are safe, survivable, and effective. Moreover, service and acquisition officials have argued that DOT&E does not have the independent authority to require and approve service-conducted follow-on operational test and evaluation after full-rate production begins. In addition, the Office is not currently required to report on whether new systems are both operationally suitable and effective before they are fielded.

DOT&E management must balance its oversight responsibilities for operational testing with the broader acquisition priorities of program managers and service test agencies. Though supportive of the Office's mission and independence, program and service representatives frequently considered the time, expense, and resources expended to

⁷Aspects of realism can include (1) equipment and personnel placed under realistic stress and operational tempo, (2) threat-representative forces, (3) end-to-end testing, (4) realistic combat tactics, (5) operationally realistic environments and targets, (6) countermeasured environments, (7) interfacing systems, (8) terrain and environmental conditions, and (9) contractor involvement.

accommodate DOT&E concerns to be ill-advised. Service officials contended that the additional testing requested by DOT&E was either unnecessary for determining the operational effectiveness or suitability of a program or unrealistic in light of the limitations in the services' testing resources.

DOT&E must manage multiple oversight, advisory, and coordination responsibilities. Several current trends may challenge DOT&E's ability to manage its workload and its ability to impact operational test and evaluation. These trends include (1) service challenges to DOT&E's authority to require and oversee follow-on operational testing and evaluation, (2) a decline in resources available for oversight, (3) an expansion of DOT&E involvement in activities other than oversight of major acquisition programs, (4) participation of DOT&E in the acquisition process as a member of working-level integrated product teams, and (5) greater integration of developmental and operational testing. These trends make it imperative that DOT&E prioritize its workload to achieve a balance between the oversight of major defense acquisition programs and other initiatives important to the quality of operational test and evaluation.

DOT&E Advocates More Thorough Testing Than the Services

A frequent complaint among representatives of the services' operational testing agencies was that DOT&E frequently demanded more tests than were proposed by the operational test agencies in draft master plans or test plans. Operational test agency representatives contended that the additional testing was either unnecessary for determining the operational effectiveness or suitability of a program or unrealistic in light of the limitations in the services' testing resources. However, our review indicated that DOT&E urged more testing to reduce the level of risk and number of unknowns prior to the decision to begin full production, while program and service officials typically sought less testing and were willing to accept greater risk when making production decisions. The additional testing DOT&E advocated, often over the objections of service testers, served to meet the underlying objectives of operational testing—to reduce the uncertainty and risk that systems entering full-rate production would not fulfill their requirements.

The impact of DOT&E oversight varies with the system under development. Table 1 summarizes the types of impacts that DOT&E advocated or facilitated in operational testing among the 13 cases we studied. While the impacts vary, one consistent pattern in our case studies was a reduction in uncertainty regarding the weapon systems' suitability or effectiveness

prior to the full-rate production decision. Each of the impacts are discussed in more detail in tables 2-6 and in subsequent sections.

Table 1: Types of Impacts on the Operational Testing of 13 Systems Due to DOT&E Oversight

System	More testing advocated and conducted	More realism included in test design	Enhancements made in data collection or analysis	DOT&E's conclusion deviated from the service's	Follow-on operational test and evaluation advocated and planned or conducted
AH-64D Longbow Apache helicopter	X	X	X	X	X
ASPJ ^a jammer	X		X		
C-17A aircraft	X				X
E-3 AWACS ^b (RSIP ^c)		X	X		
F-22 fighter	X	X			X
Javelin missile	X	X		X	
Joint STARS ^d		X	X	X	X
LPD-17 assault ship	X		X		
M1A2 tank			X	X	X
Sensor fuzed weapon	X	X		X	X
Standard missile	X	X			
Tomahawk Weapon System		X	X		
V-22 aircraft		X			

Note: The absence of an "X" does not necessarily indicate the absence of DOT&E impact. For example, blanks may occur where DOT&E and the service agreed on issues; however, the deterrent effect of DOT&E oversight is unquantifiable. In addition, blanks may occur because the system has not yet progressed through the entire acquisition process.

^aAirborne Self-Protection Jammer.

^bAirborne Warning and Control System.

^cRadar System Improvement Program.

^dSurveillance Target Attack Radar System.

DOT&E Oversight Led to More Testing Than Proposed by the Operational Test Agencies

Two of DOT&E's typical concerns in reviewing service test plans are that the proposed test methodologies enable (1) comparisons of a system's effectiveness through side-by-side testing between the existing and modified systems and (2) assessments of a system's reliability through a sufficient number of test repetitions. Table 2 illustrates examples of cases

where additional testing was conducted at DOT&E's insistence or with DOT&E's support to alleviate these and other types of effectiveness and suitability concerns.

Table 2: Examples of Programs That Expanded Testing Due to DOT&E Oversight

System	Expanded testing	Impact
AH-64D Longbow Apache helicopter	DOT&E insisted that the Army include a baseline AH-64A company in gunnery and force-on-force exercises to ensure direct comparability with the Longbow.	Testers were able to demonstrate the gunnery performance improvements of the AH-64D. These improvements included that (1) the AH-64D had 300 instances of lethality compared to 75 for the AH-64A, (2) the AH-64D was approximately 8 times more survivable than the AH-64A, and (3) the AH-64D had zero fratricide instances compared to 34 for the AH-64A.
ASPJ jammer	In follow-on operational test and evaluation of the F-14D begun in 1995, DOT&E insisted that the scope of the test plan address the ASPJ's contribution to the aircraft's survivability—not merely the jammer's compatibility with the aircraft's avionics. This expansion of the scope necessitated an additional 18 open air flight tests to measure the ASPJ's effectiveness against air-to-air threats and a requirement to gather suitability data pertaining to ASPJ, including its built-in test equipment. ^a	The revised test plan enabled testers to address the critical operating issue—that the F-14D is more survivable with the ASPJ as part of its electronic warfare suite than without it.
C-17 aircraft	The ability to safely perform a mass personnel airdrop while flying in close formation is a key Air Force capability needed to conduct a strategic brigade airdrop. DOT&E insisted that an airdrop of a brigade slice of personnel and equipment be done. The Air Force's position was that the airdrop was unnecessary before the full-rate production decision and that the use of the aircraft in airdrops would be determined after the full-rate production decision.	DOT&E forced testing that confirmed operational limitations, and the Army has yet to approve mass airdrops of personnel from C-17s flying in close formation. Operational tests identified specific problems with the C-17's airdrop capability—that with the air turbulence created in the wake of the aircraft, flying in close formation can cause the parachutes dropping from aircraft to oscillate, partially deflate, or collapse. These conditions could result in serious injury or death to paratroopers.
F-22 fighter	DOT&E and the Air Force agreed to a balanced approach of open-air testing, full mission simulation, and digital models against then-current and future threats in an overall F-22 and F-15 effectiveness analysis.	The use of multiple testing and evaluation techniques will reduce uncertainty in system effectiveness more than the Air Force's initial preference to use test results to support evaluation by modeling.
Javelin missile	DOT&E insisted that the system undergo additional operational testing prior to the full-rate production decision in 1997 because over 50 design changes had been made to the system since initial operational test and evaluation in 1993. The Army claimed that successful passage of technical tests was adequate assurance of suitability for combat and did not originally intend to conduct operational tests until 1998, over a year after the start of full-rate production. ^b	The test provided additional confidence that the weapon system's modifications had not affected Javelin's suitability for combat.

(continued)

System	Expanded testing	Impact
Javelin missile (con't)	Based on data collected from initial operational testing, DOT&E disagreed with the Army's conclusion that the Javelin was suitable for combat and supported the Army's operational test agency in requiring the program manager to conduct an operational test to confirm the unit's reliability.	Before the additional test was conducted, the Army modified components of the command launch unit to increase its reliability. The subsequent test demonstrated that the modifications were successful. The test also provided two additional benefits. First, missile failures during the test led to discovery and correction of a design flaw that prevented the missiles from leaving the launch tube when the gunner pulled the trigger. Second, while developing the test plan, DOT&E discovered that the Army had no Javelin-specific tactical doctrine and recommended the Army study this deficiency. As a result, the Army developed operational tactics to guide officers in integrating Javelin with other antitank systems.
LPD-17 assault ship	The originally proposed operational test for the LPD-17 consisted of at-sea steaming and some landing craft air cushion (LCAC) operations. DOT&E forced the incorporation of full-scale assault operations with LCACs, aircraft, ground assault equipment, and personnel.	The expanded scope of the test plan will more closely encompass the range of system requirements as well as enhance the realism of the test scenario.
Sensor fuzed weapon	DOT&E insisted on a second phase of operational test and evaluation before the full-rate production decision that the Air Force did not want to conduct.	The additional testing of system issues not fully tested in the first phase (such as additional countermeasures, multiple releases, and an alternate target formation) reduced uncertainty in system effectiveness and reliability.
Standard missile SM-2	DOT&E insisted on and obtained five flight tests of the User Operational Evaluation System SM-2 block IVA missile, a theater ballistic missile defense system. The Navy planned only two at-sea safety flights against nonthreat-representative targets. Some of the new flight tests will be conducted against threat-representative targets from the integrated AEGIS system.	DOT&E's insistence on additional testing has lowered the technical risk of the program by providing for a series of tests to establish system level validation. These tests will help to demonstrate the level of reliability and effectiveness of the SM-2 block IVA missile.

^aSee Electronic Warfare (GAO/NSIAD-96-109R, Mar. 1, 1996).

^bSee Army Acquisition: Javelin Is Not Ready for Multiyear Procurement (GAO/NSIAD-96-199, Sept. 26, 1996).

DOT&E Oversight Led to More Realistic Testing Than Proposed by the Operational Test Agencies

Table 3 illustrates examples where the design or conduct of operational testing was modified at DOT&E's insistence or with DOT&E's support to increase the realism of test conditions and reduce the uncertainty of system suitability or effectiveness.

Table 3: Examples of Programs That Conducted More Realistic Testing Due to DOT&E Oversight

System	Enhanced realism in tests	Impact
AH-64D Longbow Apache helicopter	DOT&E required a demanding air defense network, directly intervening to ensure that a specific threat would be present in the force-on-force trials.	The testing revealed operational limitations of the AH-64D variant without the fire control radar and thereby raised the issue of the appropriate mix of variants to procure. The AH-64D variant with the fire control radar was unable to reduce the air defense threat sufficiently to allow the variant without the fire control radar to move into battle positions without significant possibility of being engaged by those air defense units.
E-3 AWACS (RSIP)	DOT&E insisted that (1) mission crews comprise a cross section of typical AWACS aircrew members, (2) RSIP be employed against an array of actual Soviet and other threats, and (3) the system be used in eight different terrain combinations in both the United States and Europe.	Reduced uncertainty of system effectiveness because (1) AWACS personnel from the engineering and developmental test sorties were excluded, resulting in the use of two test crews comprised of a typical ratio of U.S. and Canadian deployment personnel and (2) actual threats and realistic environments were incorporated.
F-22 fighter	DOT&E was instrumental in ensuring that a full mission simulator was developed for comparison testing using validated software and hardware, insisting that the functionality and fidelity of the simulation be validated by open air flight data. DOT&E insisted that the test and evaluation master plan include high tempo demonstrations to test the required sortie generation rate.	The credibility of the full mission simulator (used to compare relative mission effectiveness of the F-15 and F-22) will be enhanced. The confidence level of the model's prediction is enhanced by introducing surge data from actual operations.
Javelin missile	DOT&E required Army troops to carry the missile a representative distance during missions and prior to actual firings to ensure that the missile's reliability would not be affected by field handling.	The Army found that missiles carried during the test failed to leave the launch tube because of a faulty design of the external restraining pin-wiring harness. This finding led the Army to redesign the assembly, which prevented potential missile malfunctions in combat situations.
Joint STARS	In the development of the test plan, DOT&E encouraged participation of Air Force and Army testers in training exercises at the National Training Center as a way to enhance test realism.	Deployment of the system to Bosnia precluded testing at the National Training Center, but the test design precedent was established.

(continued)

System	Enhanced realism in tests	Impact
Sensor fuzed weapon	During the second phase of initial operational test and evaluation, DOT&E required an extensive validation of the infrared signature and the use of countermeasures, insisted on all-weather and all-altitude testing at numerous test sites; insisted on realistic and comprehensive countermeasures testing; and ensured realistic targets were made available for testing.	The enhanced realism of testing reduced uncertainty of system effectiveness at low altitudes and confirmed decreased effectiveness as altitude, dive angle, and time of flight increase.
Standard missile SM-2	During the review of the Navy's draft test and evaluation master plan for the SM-2 block IV, DOT&E identified inadequacies in aerial target programs and required that threat-representative targets be available for operational testing.	The need for realistic aerial targets is a significant issue cutting across all Navy surface antiair warfare programs such as the Phalanx Close-In Weapon System and the Rolling Airframe Missile, as well as the various SM-2 blocks.
Tomahawk Weapon System	DOT&E was instrumental in ensuring that only ship crews were used during the testing of the all-up-rounds ^a and the Tomahawk Weapon Control System. Support personnel conducted testing, while contract personnel maintained the equipment as they do in actual operations.	The use of realistic operators reduced uncertainty in system reliability and effectiveness.
V-22 aircraft	DOT&E has emphasized the effects of the V-22 downwash on personnel and material in the vicinity of the hovering aircraft and the need to test in more realistic ship and landing zone environments.	The test program has been revised to conduct downwash testing in 1997 rather than 1999 to address the concerns of DOT&E and others.

^aEach Tomahawk missile variant is contained within a pressurized canister to form an all-up-round.

DOT&E Oversight Led to Changes in the Data Analysis Plan

DOT&E can insist on or support changes in data analysis plans that provide more meaningful analyses for decisionmakers. Table 4 illustrates instances in which DOT&E altered the proposed data collection or analysis plans to enhance the reliability or utility of the test data.

Table 4: Examples of Programs in Which Changes Were Made in the Data Analysis Plan Due to DOT&E Oversight

System	Changes in data analysis plan	Impact
AH-64D Longbow Apache helicopter	DOT&E insisted on performance criteria to assess the superiority of the AH-64D over the AH-64A. The criteria—a 20-percent improvement—had not formally been included in the test and evaluation master plan. DOT&E required measures that addressed the number of targets killed and helicopters lost.	DOT&E input allowed testers to more accurately compare the AH-64D to the AH-64A in quantifiable categories of lethality, survivability, and fratricide.
ASPJ jammer	DOT&E required the Navy to test the ASPJ against the type of missile that shot down an F-16 over Bosnia in June 1995. ^a	The Navy determined that the ASPJ was effective against that threat.
	DOT&E was instrumental in establishing a requirement to gather suitability data on its built-in test equipment. While the contractor reported improvement in previously unreliable built-in test equipment, DOT&E questioned the data collection and interpretation.	Independent oversight of ASPJ's suitability assessment confirmed ongoing concerns with system reliability.
E-3 AWACS (RSIP)	DOT&E insisted that service personnel be trained to operate contractor data extraction systems, thereby removing the contractor from the process and ensuring data integrity. DOT&E reviewed a major radar failure and discovered an error in the technical path described by the service.	Reduced uncertainty of system effectiveness because the contractor was removed from data processing ensuring test integrity.
Joint STARS	DOT&E insisted that the Air Force modify its original technical requirements to include measures of effectiveness that directly addressed the missions of surveillance, targeting, and battlement management. DOT&E stressed differentiation between user and system requirements.	The change in test measures resulted in test data that were more operationally relevant to system effectiveness.
LPD-17 assault ship	DOT&E insisted on measures of effectiveness that addressed the movement of men and equipment ashore rather than the Navy's original requirements that focused on technical specifications.	The change in test measures will result in test data that are more operationally relevant to system effectiveness.

(continued)

System	Changes in data analysis plan	Impact
M1A2 tank	DOT&E required that the Army use credible data for the determination of reliability in follow-on operational test and evaluation. The Army proposed the use of failures and other secondary measures that would not provide a credible basis for reversing the results of initial operational test and evaluation. DOT&E insisted that the operational testing be conducted to compare the M1A2 with the M1A1. Several improvements in the M1A2 addressed command and control that could not be directly measured. By conducting several operations with both tanks, the difference in movements and coordination could be examined to determine the value of the command and control improvements. By adding uncertainty to test scenarios, DOT&E enabled the Army operational test agency a means to identify differences between the M1A1 and M1A2 models.	Reduced uncertainty of improved effectiveness and suitability of the M1A2 compared with the M1A1.
Tomahawk Weapon System	DOT&E was instrumental in ensuring that the effectiveness of mission planning systems was validated using high-fidelity models and simulations and that bit-by-bit checks were conducted to validate the effectiveness of functional operations of the planning system.	More rigorous data collection and validation reduced uncertainty of system effectiveness.

^aSee Airborne Self-Protection Jammer (GAO/NSIAD-97-46R, Jan. 29, 1997).

DOT&E Interpreted the Results of Some Testing Less Favorably Than the Operational Test Agencies

DOT&E's independent analysis of service test data may confirm or dispute the results and conclusions reported by the service. In the cases described in table 5, DOT&E's analysis of service operational test and evaluation data resulted in divergent, often less favorable conclusions than those reached by the service.

Table 5: Examples of Programs in Which DOT&E and Service Conclusions Differed

System	Conflicting test results	Impact
AH-64D Longbow Apache helicopter	DOT&E's independent analysis of the test data identified a predominant firing technique that had not previously been identified as useful. Though the technique was never anticipated to be used so extensively and had not been considered in the development of the Longbow's tactics, techniques, and procedures, DOT&E determined that over half of the operational test engagements were conducted using this technique. Nonetheless, this revelation was not in the Army test report.	The Army will conduct a series of simulations and additional missile firings to determine the factors affecting the overall effectiveness of the technique and its relative effectiveness to the primary modes of engagement, thereby increasing certainty in system effectiveness.
Javelin missile	DOT&E did not use reliability data from the pre-initial operational test and evaluation period because the data were not realistic; as a result, DOT&E found the command launch unit failed to meet its reliability criteria, differing from the Army's report.	The Army made numerous design changes to the launch unit and round before the contractor initiated low-rate production.
Joint STARS	DOT&E disagreed with the Air Force operational test agency's positive assessment of the operational suitability and effectiveness of Joint STARS following its deployment to Operation Joint Endeavor. DOT&E concluded that Joint STARS met one of three critical operational effectiveness issues—with limitations, while the other two effectiveness issues could not be determined. Overall, the Air Force's conclusion was "suitable with deficiencies"; DOT&E's conclusion was "as tested is unsuitable." ^a DOT&E and the Air Force operational test agency also disagreed on how to report data when terrain masking occurred. DOT&E objected to the Air Force's phrasing "nothing significant to report," when in fact nothing could be seen.	DOT&E's Beyond-LRIP report indicated not only the Joint STARS' disappointing test results but also the need for extensive follow-on operational test and evaluation. Subsequently, the Joint STARS acquisition decision memorandum required that the test and evaluation master plan be updated and that follow-on operational test and evaluation address the deficiencies identified in initial operational test and evaluation by DOT&E.
M1A2 tank	DOT&E evaluated the tank as not operationally suitable—a finding at odds with Army testers. DOT&E determined that the tank was unreliable and unsafe due to uncommanded turret movements, hot surfaces that caused contact burns, and inadvertent firing of the .50 caliber machine gun.	Follow-on operational test and evaluation was conducted to determine if the Army's design changes had improved the system. The suitability problems persisted and the follow-on operational test and evaluation was suspended. New design changes were made and a second follow-on operational test and evaluation was conducted, which determined that the safety issues were resolved and that the tank is now operationally suitable.
Sensor fuzed weapon	Based on the results of the first phase of operational test and evaluation ending in 1992, the Air Force concluded that the sensor fuzed weapon was "suitable and effective for combat." In contrast, DOT&E concluded from the same tests that the system was only "potentially operationally effective and suitable."	As a result of the unresolved issues in 1992, a second phase of operational test and evaluation was planned and executed, leading DOT&E to conclude in 1996 that the system was operationally suitable and effective—when employed at low altitude using level or shallow angle dive deliveries.

(Table notes on next page)

^aSee Tactical Intelligence: Joint STARS Full-Rate Production Decision Was Premature and Risky (GAO/NSIAD-97-68, Apr. 25, 1997).

DOT&E Recommended Follow-On Operational Test and Evaluation

When DOT&E concludes that a weapon system has not fully demonstrated operational suitability or effectiveness, or if new testing issues arise during initial operational test and evaluation, it may recommend that follow-on operational test and evaluation be done after the full-rate production decision. Table 6 identifies follow-on operational test and evaluation that DOT&E supported.

Table 6: Examples of Programs in Which DOT&E Called for Follow-On Operational Test and Evaluation

System	Advocated follow-on operational test and evaluation	Impact
AH-64D Longbow Apache helicopter	DOT&E sought follow-on operational test and evaluation to characterize the Hellfire missile's performance when using lock-on before launch-inhibit technique. This method of engagement enables crews to immediately take cover after target detection and fire at moving targets from those covered locations. This method was used in over half of the operational test engagements, though it had not been considered sufficiently significant to incorporate in the Longbow's tactics, techniques, and procedures.	The use of this technique was not fully anticipated prior to initial operational test and evaluation. Its use provided an unexpected level of survivability for the AH-64D crews. This technique had been subjected to little, if any, developmental testing. Further testing will establish its probability of hit. The Army operational test agency plans to fire 8 to 10 missiles in August 1998.
C-17A aircraft	DOT&E urged follow-on operational test and evaluation to demonstrate the system's ability to meet operational readiness objectives, including combination and brigade airdrops, and software maturity.	The Air Force has undertaken further testing with the Army to overcome system deficiencies and demonstrate effectiveness. The Army is formulating a time requirement of about 30 minutes for completing a strategic airdrop. The C-17 currently has a 5.5 minute aircraft separation restriction that essentially prohibits formation flying and therefore requires 2.5 hours to complete a strategic airdrop. This resulted in continuing efforts to resolve these operational limitations.
F-22 aircraft	DOT&E insisted that the test and evaluation master plan require follow-on operational test and evaluation on two capabilities that will not be released until after initial operational test and evaluation: employment of the Joint Direct Attack Munition and Cruise Missile Defense.	The commitment to test these capabilities is formally acknowledged.
Joint STARS	DOT&E stated in its Joint STARS Beyond-LRIP report that only 18 of 71 performance criteria tested were demonstrated by the system and that further testing was required for the remaining 53.	The Joint STARS acquisition decision memorandum directed additional testing to address suitability deficiencies in logistics and software.

(continued)

System	Advocated follow-on operational test and evaluation	Impact
M1A2 tank	The M1A2, during initial operational test and evaluation in 1993, failed to meet the combat mission reliability threshold, encountered an excessive number of battery failures, consumed 15 percent more fuel, exhibited uncommanded main gun/turret movements and inadvertent .50 caliber machine-gun firing that made the tank unsafe. DOT&E, through a Secretary of Defense letter accompanying the M1A2 Beyond-LRIP report to Congress, required follow-on operational test and evaluation on M1A2 suitability issues when the Army claimed it was unnecessary.	The Army executed a program to correct the deficiencies identified during initial operational test and evaluation and conducted follow-on operational test and evaluation in 1995. Suitability issues, such as uncommanded turret movement and power loss, were again experienced. The follow-on operational test and evaluation was put on hold until additional corrective actions could be applied. Follow-on operational test and evaluation resumed in July 1996. The safety problems were found to have been addressed by the design changes, and there were no observed instances of the problems experienced during initial or beginning follow-on operational test and evaluation.
Sensor fuzed weapon	The test and evaluation master plan for the second phase of operational test and evaluation specified a series of follow-on operational test and evaluations that would address how well the addition of the Wind Compensated Munition Dispenser and the preplanned product improvements will rectify system limitations.	Follow-on operational test and evaluation ensures further investigation of system limitations known at the time of the full-rate production decision.

Strengths and Weaknesses of Current Organizational Framework

The existence of a healthy difference of opinion between DOT&E and the acquisition community is a viable sign of robust oversight. In nearly all of the cases we reviewed, the services and DOT&E cited at least one testing controversy. For example, services differ on how they view the relationship between operational testing and their development of tactics, techniques, and procedures. In addition, DOT&E's ability to independently view the development and testing of new systems across the services brings value to the context of testing. However, several current trends have the potential to adversely affect DOT&E's independence and its ability to affect operational test and evaluation, including (1) service challenges to DOT&E's authority to require and oversee follow-on operational testing and evaluation, (2) declining resources available for oversight, (3) the management of limited resources to address competing priorities, (4) DOT&E's participation in the acquisition process as a member of the program manager's working-level integrated product teams, and (5) greater integration of developmental and operational testing. DOT&E's impact on operational testing is dependent upon its ability to manage these divergent forces while maintaining its independence.

Independence Is the Key to DOT&E's Effectiveness

Although the acquisition community has three central objectives—performance, cost, and schedule—DOT&E has but one: operational testing of performance. These distinct priorities lead to testing disputes. Characteristically, the disputes for each system we reviewed revolved around questions of how, how much, and when to conduct operational testing, not whether to conduct operational testing. Conflicts encompassed issues such as (1) how many and what types of tests to conduct; (2) when testing should occur; (3) what data to collect, how to collect it, and how best to analyze it; and (4) what conclusions were supportable, given the analysis and limitations of the test program. The foundation of most disputes lay in different notions of the costs and benefits of testing and the levels of risk that were acceptable when making full-rate production decisions. DOT&E consistently urged more testing (and consequently more time, resources, and cost) to reduce the level of risk and number of unknowns before the decision to proceed to full-rate production, while the services consistently sought less testing and accepted more risk when making production decisions. Among our case studies, these divergent dispositions frequently led to healthy debates about the optimal test program, and in a small number of cases, the differences led to contentious working relations.

In reviews of individual weapon systems, we have consistently found that testing and evaluation is generally viewed by the acquisition community as a requirement imposed by outsiders rather than a management tool to identify, evaluate, and reduce risks, and therefore a means to more successful programs. Developers are frustrated by the delays and expense imposed on their programs by what they perceive as overzealous testers. The program office strives to get the program into production despite uncertainties that the system will work as promised or intended. Therefore, reducing troublesome parts of the acquisition process—such as operational testing—is viewed as a means to reduce the time required to enter production.

Nonetheless, the commanders and action officers within the service operational test agencies were nearly unanimous in their support for an independent test and evaluation office within OSD. For example, the Commander of the Army's Operational Test and Evaluation Command commended the style and orientation of the current DOT&E Director and affirmed the long-term importance of the office and its independent reporting responsibilities to Congress. The Commander of the Navy's Operational Test and Evaluation Force stated that the independence of both DOT&E and the operational test agency was an essential element in

achieving their common goal of ensuring that new programs pass sufficiently rigorous and realistic operational testing prior to fielding. The Commander of the Air Force's Operational Test and Evaluation, while critical of DOT&E oversight of several major weapon systems, said that the services were well served by DOT&E's potential to independently report to Congress. Moreover, nearly all the operational test agency action officers we interviewed participate in the integrated product teams with the DOT&E action officers and recognized the value of the Office's independent oversight role. The action officers within the service testing organizations also have a degree of independence that enables them to represent the future users of systems developed in the acquisition community. These action officers stated that their ability to voice positions unpopular with the acquisition community was strengthened when DOT&E separately supported their views.

In discussions with over three dozen action officers and analysts responsible for the 13 cases we reviewed, the independence of DOT&E emerged as the fundamental condition to enable effective and efficient oversight. The foundation of interagency (i.e., DOT&E and service operational test agencies) relations is based on the independence of DOT&E, its legislative mandate, and its independent reporting to Congress. DOT&E is outside the chain of command of those responsible for developing and testing new systems. The services need to cooperate with DOT&E primarily because the Office must approve all test and evaluation master plans and operational test plans. Moreover, DOT&E independently reports on the operational suitability and effectiveness at a system's full-rate production milestone, a report that is sent separately to Congress.

Unfavorable Reports on Operational Testing Do Not Always Inhibit Full-Rate Production

DOT&E's report on a system's operational suitability and effectiveness is only one of several inputs considered before the full-rate production decision is made. An unfavorable DOT&E report does not necessarily prevent full-rate production. In each of the cases cited below, an affirmative full-rate production decision was made despite a DOT&E report concluding that the system had not demonstrated during operational test and evaluation that it was both operationally suitable and operationally effective:

- Full-rate production of the M1A2 tank was approved despite DOT&E's report that found the system unsuitable.
- Full-rate production of Joint STARS was approved, though the system demonstrated only limited effectiveness for "operations other than war"

and found "as tested is unsuitable." Only 18 of the 71 performance criteria were met; 53 others required more testing.

- Full-rate production of the C-17 Airlifter was approved despite a number of operational test and evaluation deficiencies, including immature software and failure to meet combination and brigade airdrop objectives.

Services Contest DOT&E Oversight of Follow-on Operational Test and Evaluation

The services contend that DOT&E does not have authority to insist on, or independently approve the conduct of, follow-on operational test and evaluation. However, in several of the systems we reviewed, DOT&E overcame service opposition and monitored follow-on operational test and evaluation. It used several means to achieve success, such as (1) incorporating follow-on operational test and evaluation in test and evaluation master plans developed and approved prior to the full-rate production decision milestone; (2) persuading the Secretary of Defense to specify follow-on operational test and evaluation, and DOT&E's oversight role, in the full-rate production acquisition decision memorandum; and (3) citing policy, based on title 10, that entitles DOT&E to oversee operational test and evaluation whenever it occurs in the acquisition process.⁸

Nonetheless, DOT&E action officers stated that the service's acceptance of DOT&E's role in follow-on operational test and evaluation varies over time, by service and acquisition system, and is largely dependent upon the convictions of executives in both the services and DOT&E. Among the cases reviewed in this report, the services offered a variety of arguments against DOT&E's having a role in follow-on operational test and evaluation. They specifically asserted the following:

- DOT&E need not be involved because the scope of follow-on operational test and evaluation is frequently less encompassing than initial operational test and evaluation. Follow-on operational test and evaluation has been characterized as testing by the user to determine the strengths and

⁸In March 1997 DOT&E issued the "Policy on DOT&E Oversight of Systems in Follow-on Operational Test and Evaluation (FOT&E)." The Director stated that 10 U.S.C. section 189 provides DOT&E with the authority to oversee follow-on operational test and evaluation. Specifically, DOT&E shall oversee follow-on operational test and evaluation to (1) refine estimates made during operational test and evaluation, (2) complete initial operational test and evaluation activity, (3) verify correction of deficiencies, (4) evaluate significant changes to design or employment, and (5) evaluate the system to ensure that it continues to meet operational needs and retains effectiveness in a substantially new environment or against a new threat. The Director elaborated by specifying that normal DOD 5000.2R documental and approval requirements apply.

weaknesses of the system and to determine ways to compensate for, or fix, shortcomings observed in initial operational test and evaluation.⁹

- Title 10 provides DOT&E with the authority to monitor and review—but not necessarily approve—service follow-on operational test and evaluation plans.¹⁰
- Follow-on operational test and evaluation is unnecessary when a system is found to be operationally effective and suitable during initial operational test and evaluation—even though DOT&E does not concur.¹¹

A clear distinction between DOT&E oversight in follow-on operational test and evaluation versus initial operational test and evaluation is that DOT&E is not required to report follow-on operational test and evaluation results to Congress in the detailed manner of the Beyond-LRIP report. Therefore, even if follow-on operational test and evaluation is conducted to assess modifications to correct effectiveness or suitability shortcomings reported to Congress in the Beyond-LRIP report, there is no requirement that Congress receive a detailed accounting of the impact of these modifications.

DOT&E's Resources Are Declining

DOT&E's primary asset to conduct oversight—its cadre of action officers—has decreased in size throughout the decade. This creates a management challenge for the Office because at the same time staff has decreased, the number of programs overseen by DOT&E has increased. As illustrated in table 7, authorized staffing declined from 48 in fiscal year 1990 to 41 in fiscal year 1997, as did funding (in constant dollars) from \$12,725,000 in fiscal year 1990 to \$11,437,000 in fiscal year 1997. The decline in DOT&E funding is consistent with the general decline in DOD appropriations during this period. However, since fiscal year 1990, while the authorized staffing to oversee operational test and evaluation has declined by 14.6 percent, the number of systems on the oversight list has increased by 17.7 percent.

⁹In the case of Joint STARS, the acquisition decision memorandum required the Air Force and the Army to update a test and evaluation master plan for OSD approval—but did not require DOT&E approval. Moreover, the Director of Air Force Test and Evaluation termed post-milestone III testing as “regression testing” and emphasized that DOT&E had no oversight role.

¹⁰In two case study systems, the C-17 and Joint STARS, the Air Force provided DOT&E with a copy of its follow-on operational test and evaluation test plans for review but did not allow sufficient time and had no expectation that DOT&E would approve the plans prior to the initiation of testing.

¹¹The acquisition decision memorandum for the M1A2 tank required the Army to conduct follow-on operational test and evaluation (with DOT&E oversight) on safety and suitability shortcomings identified by DOT&E in initial operational test and evaluation, though the Army had already determined that the system was operationally suitable as tested.

Table 7: DOT&E Staffing and Funding

Dollars in thousands

	Fiscal year							
	1990	1991	1992	1993	1994	1995	1996	1997
Funding ^a	\$12,725	\$13,550	\$12,836	\$12,333	\$11,450	\$12,501	\$12,183	\$11,437
Authorized staffing	48	46	44	44	43	43 ^b	42	41
Oversight programs	186	207	204	191	199	202	219	219

^aFunding for operational test and evaluation program element only; funding provided for the live fire test and evaluation program element assumed by DOT&E beginning in fiscal year 1995 is not reflected in the funding data for fiscal years 1995-97.

^bThe authorized end strength for DOT&E beginning in fiscal year 1995 increased by four as a result of the congressionally directed (Federal Acquisition Streamlining Act of 1994, P.L. 103-355) move of live fire test and evaluation responsibilities to DOT&E. Since these positions are dedicated to live fire testing and not operational testing, their numbers are not reflected in this table.

DOT&E's Limited Resources Must Address Competing Priorities

With declining resources, DOT&E must manage competing priorities related to its oversight, advisory, and coordination responsibilities. DOT&E must balance the continuing need to allocate resources to these different priorities while not being perceived as having lost any independence. DOT&E management has flexibility in defining some portion of the scope of its oversight and has continued to electively oversee a substantial number of nonmajor defense acquisition programs and assumed a leading role in advocating an examination of the modernization needs of the test and evaluation infrastructure.

DOT&E Continues to Oversee a Substantial Number of Nonmajor Programs

Between fiscal year 1990 and 1996, the number of nonmajor acquisition programs overseen annually by DOT&E ranged between 19 and 43. In fiscal year 1996, when the oversight list reached a peak of 219, 1 of every 8 programs was listed at the discretion of DOT&E. Thus, during this period when the resources to oversee operational testing declined and acquisition reforms have placed additional burdens on oversight staff, the directors of DOT&E continued to place extra responsibility on their staff by augmenting the required oversight of major acquisition programs with a substantial number of optional systems.

Despite a relative decline in resources for oversight, DOT&E management has also elected to assume "a larger role in test resource management planning and leadership in an attempt to achieve much-needed resource modernization."¹² Although the Director is designated as the principal

¹²Director, Operational Test and Evaluation, FY95 Report, March 1996, p. I-6.

adviser to the Secretary of Defense and the Under Secretary of Defense for Acquisition and Technology on operational test and evaluation, including operational test facilities and equipment,¹³ assuming the larger role defined by DOT&E may be at the expense of its testing oversight mission and perception of independence. The DOT&E Director is now an adviser to the Central Test and Evaluation Investment Program and previously served as Chairman of the Test and Evaluation Committee. The Committee is responsible for the investment program and presides over the planning, programming, and budgeting for development and operational test resources. When the Director served as chairman, we questioned whether these ties created the perception that the Director was not independent from developmental testing.¹⁴ This issue may resurface as DOT&E seeks a larger role in test resource management planning. Also, as the emphasis, cost, and time for operational test and evaluation are increasingly questioned in the drive to streamline acquisition, and as oversight assets are stretched, new DOT&E initiatives may stress the Office's capacity to manage oversight effectively.

DOT&E Participation in Working-Level Integrated Product Teams Has the Potential to Complicate Independence

In May 1995, the Secretary of Defense directed DOD to apply the integrated product and process development concept—using integrated product teams—throughout the acquisition process. The revised DOD acquisition regulations (DOD 5000.2-R March 1996) also addressed the use of empowered integrated product teams at the program office level. DOT&E action officers participate as members of the working-level integrated product teams, and the DOT&E Director is a member of the overarching team. One objective of integrated product teams, and DOT&E participation in particular, is to expedite the approval process of test documents by reaching agreement on the strategy and plan through the identification and resolution of issues early, understanding the issues, and documenting a quality test and evaluation master plan that is acceptable to all organizational levels the first time. Integrated product teams are designed to replace a previously sequential test and evaluation master plan development and approval process and therefore enhance timeliness. While this management tool could increase communication between

¹³10 U.S.C. section 139 assigns six responsibilities to the Director, the fifth of which is to "review and make recommendations to the Secretary of Defense on all budgetary and financial matters relating to operational test and evaluation, including operational test facilities and equipment {emphasis added}, in the Department of Defense."

¹⁴In Test and Evaluation: The Director, Operational Test and Evaluation's Role in Test Resources (GAO/NSIAD-90-128, Aug. 27, 1990), we found that the Director's independence was jeopardized because the Director had influence over the types of development test assets used by the services. Responsibility for developmental test resources rests with the services. In 1987 Congress amended DOT&E's statute to emphasize the separation of operational testing from functions associated with developmental testing by stating that "the Director may not be assigned any responsibility for developmental test and evaluation, other than the provision of advice to officials responsible for such testing."

testers and the program managers, it also poses a challenge to DOT&E independence. The challenge was recognized by the Department of Defense Inspector General (DOD IG) when after reviewing the conduct of operational testing it subsequently recommended that "to meet the intent of 10 U.S.C. 139, DOT&E should be a nonvoting member [of the working-level integrated product team] so as to maintain his independence."¹⁵ {emphasis added} Though integrated product teams were not used throughout the entire time period covered by this report, several action officers noted that this management tool created threats to their effectiveness other than having their positions out-voted. One DOT&E action officer reported having the lone dissenting opinion in a meeting of 30 participants seeking to reach consensus and resolve issues early. The pressure of maintaining independent, contrary positions in large working groups can be a test. Several DOT&E representatives also noted that the frequency of integrated product team meetings to cover the multiple systems for which they were responsible made it impossible for them to attend all, thereby lessening the possibility that testing issues can be identified and resolved as early as possible.

Moreover, program managers and DOT&E pursue different objectives through integrated product teams. The services and program managers view the teams as a way to facilitate their program objectives for cost, schedule, and performance; DOT&E's objective is oversight of performance through operational testing. The program managers and DOT&E share a desire to identify testing issues as early as possible. However, the priority of the program manager to resolve these issues as early as possible through the teams may conflict with DOT&E's mission. DOT&E must remain flexible and react to unknowns as they are disclosed during developmental testing, operational assessments, and initial operational test and evaluation. Thus, DOT&E's participation on the teams is a natural source of tension and a potential impediment to the team's decision-making. The challenge for DOT&E action officers is to maintain an independent and potentially contrary position in an ongoing working group during the life of a program, which may extend over several years.

Increased Integration of Developmental and Operational Testing May Attenuate Independent Oversight

The objectives of developmental and operational testing are distinct. Developmental testing determines whether a system meets its functional requirements and contractual technical performance criteria sufficiently to proceed with operational testing. Operational testing determines whether the system meets the operational requirements and will contribute to

¹⁵See Department of Defense Office of the Inspector General, Operational Testing Performed on Weapons Systems, Report No. 96-107, May 6, 1996, p. 11.

mission effectiveness in relevant operational environments sufficiently to justify proceeding with production. The integration of these two disparate test activities is proposed to save the time and resources required for testing and evaluation. The sentiment to more closely link developmental and operational testing dates from at least the 1986 Blue Ribbon Commission on Defense Management (Packard Commission), which found that "developmental and operational testing have been too divorced, the latter has been undertaken too late in the cycle, and prototypes have been used and tested far too little."¹⁶ However, both we and the DOD IG have found that systems were regularly tested before they were ready for testing. In its 1996 report, the DOD IG reported that "4 of 15 systems we examined for operational testing were not ready for testing. This situation occurred because a calendar schedule rather than system readiness often drove the start of testing."¹⁷ Similarly, we have observed numerous systems that have been pushed into low-rate initial production without sufficient testing to demonstrate that the system will work as promised or intended. Our reviews of major system development in recent years have found that because insufficient time was dedicated to initial testing, systems were produced that later experienced problems during operational testing and systems entered initial production despite experiencing problems during early operational testing.¹⁸

In 1996 the Secretary of Defense also urged the closer integration of developmental and operational testing, and combined tests where possible, in part to enhance the objectives of acquisition reform. Combined developmental and operational testing is only one of many sources of test data that DOT&E has used to foster more timely and thorough operational test and evaluation. Other sources of information include contractor developmental testing, builder's trials, component testing, production lot testing, stockpile reliability testing, and operational deployments. While DOT&E has some influence over the quality of operational testing, by independently reviewing the design, execution, analysis, and reporting of such tests, it has no direct involvement or oversight of these other sources of testing information. The use of alternative sources of test data as substitutes for operational test and evaluation will limit DOT&E's oversight mission, which was created to improve the conduct and quality of testing.

¹⁶A Quest for Excellence: Final Report by the President's Blue Ribbon Commission on Defense Management, June 1986, p. xxiii.

¹⁷Office of the Inspector General, Department of Defense, Operational Testing Performed on Weapons Systems, Report No. 96-107, May 6, 1996, p. 16.

¹⁸See Weapons Acquisition: Low-Rate Initial Production Used to Buy Weapon Systems Prematurely (GAO/NSIAD-95-18, Nov. 21, 1994).

Conclusions and Recommendations

DOT&E's challenge is to manage an expansion in independent oversight while satisfying the efficiency goals of acquisition reform and undergoing the economic pressures of downsizing. DOT&E oversight is clearly affecting the operational testing of new defense systems. DOT&E actions (such as the insistence on additional testing, more realistic testing, more rigorous data analysis, and independent assessments) are resulting in more assurance that new systems fielded to our armed forces are safe, suitable, and effective. However, DOT&E is not, by design or practice, the guarantor of effective and suitable acquisitions. DOT&E oversight reduces, but does not eliminate, the risk that new systems will not be operationally effective and suitable. Affirmative full-rate production decisions are made for systems that have yet to demonstrate their operational effectiveness or suitability. Moreover, the services question DOT&E's authority regarding follow-on test and evaluation of subsequent corrective actions by the program office.

We recommend that the Secretary of Defense revise DOD's operational test and evaluation policies in the following ways:

- Require the Under Secretary of Defense for Acquisition and Technology, in those cases where affirmative full-rate production decisions are made for major systems that have yet to demonstrate their operational effectiveness or suitability, to (1) take corrective actions to eliminate deficiencies in effectiveness or suitability and (2) conduct follow-on test and evaluation of corrective actions until the systems are determined to be operationally effective and suitable by the Director, Operational Test and Evaluation.
- Require the Director, Operational Test and Evaluation, to (1) review and approve follow-on test and evaluation master plans and specific operational test plans for major systems before operational testing related to suitability and effectiveness issues left unresolved at the full-rate production decision and (2) upon the completion of follow-on operational test and evaluation, report to Congress, the Secretary of Defense, and the Under Secretary of Defense for Acquisition and Technology whether the testing was adequate and whether the results confirmed the system is operationally suitable and effective.

Further, in light of increasing operational testing oversight commitments and to accommodate oversight of follow-on operational testing and evaluation, we recommend that the Director, Operational Test and Evaluation, prioritize his Office's workload to ensure sufficient attention is given to major defense acquisition programs.

Agency Comments and Our Evaluation

In commenting on a draft of this report, DOD concurred with our first and third recommendations and partially concurred with our second recommendation. Concerning the recommendation with which it partially concurred, DOD stated that system specific reports to the Secretary of Defense and Congress are not warranted for every system that requires follow-on operational test and evaluation. DOD pointed out that for specific programs designated for follow-on oversight, test plans are prepared to correct previously identified deficiencies by milestone III, and DOT&E includes the results of follow-on testing in its next annual report.

We continue to believe our recommendation has merit. We recommended that the Secretary require DOT&E approval of follow-on test and evaluation of corrective actions because during our review we found no consensus within the defense acquisition community concerning DOT&E's role in follow-on operational test and evaluation. In its comments DOD did not indicate whether it intended to give DOT&E a role in follow-on operational test and evaluation that is comparable to its role in initial operational test and evaluation. Moreover, we continue to believe that if a major system goes into full-rate production (even though it was deemed by DOT&E not to be operationally suitable and effective) based on the premise that corrections will be made and some follow-on operational test and evaluation will be performed, DOT&E should report, as promptly as possible, whether or not the follow-on operational test and evaluation results show that the system in question had improved sufficiently to be characterized as both operationally suitable and effective.

DOD's comments are reprinted in their entirety in appendix IV, along with our specific evaluation.

As agreed with your offices, unless you publicly announce its contents earlier, we plan no further distribution of this report until 15 days after its date of issue. We will then send copies to other congressional committees and the Secretary of Defense. We will also make copies available to others upon request.

If you have any questions or would like additional information, please do not hesitate to call me at (202) 512-3092 or the Evaluator-in-Charge, Jeff Harris, at (202) 512-3583.

A handwritten signature in black ink, appearing to read "Kwai-Cheung Chan".

Kwai-Cheung Chan
Director of Special Studies and Evaluation

Contents

Letter	1
Appendix I	28
Scope and Methodology	
Appendix II	32
Description of 13 Case Study Systems	
AH-64D Longbow Apache Helicopter	32
Airborne Self-Protection Jammer	32
C-17A Airlifter	32
E-3 AWACS Radar System Improvement Program	32
F-22 Air Superiority Fighter	32
Javelin Missile	32
Joint Surveillance Target Attack Radar System	33
LPD-17 Amphibious Assault Ship	33
M1A2 Abrams Main Battle Tank	33
Sensor Fuzed Weapon	33
Standard Missile-2	33
Tomahawk Weapon System	34
V-22 Osprey	34
Appendix III	35
DOT&E's and DOD's System Acquisition Process	
Appendix IV	37
Comments From the Department of Defense	
Related GAO Products	44
Tables	5
Table 1: Types of Impacts on the Operational Testing of 13 Systems Due to DOT&E Oversight	

Contents

Table 2: Examples of Programs that Expanded Testing Due to DOT&E Oversight	6	
Table 3: Examples of Programs That Conducted More Realistic Testing Due to DOT&E Oversight	8	
Table 4: Examples of Programs in Which Changes Were Made in the Data Analysis Plan Due to DOT&E Oversight	10	
Table 5: Examples of Programs in Which DOT&E and Service Conclusions Differed	12	
Table 6: Examples of Programs in Which DOT&E Called for Follow-On Operational Test and Evaluation	13	
Table 7: DOT&E Staffing and Funding	19	
Table I.1: Characteristics of Weapon Systems Used for Case Studies	28	
Figure	Figure III.1: DOD's Weapon System Acquisition Process	35

Abbreviations

AWACS	airborne warning and control system
DOD	Department of Defense
DOD IG	Department of Defense Inspector General
DOT&E	Office of the Director of Operational Test and Evaluation
IDA	Institute for Defense Analyses
Joint STARS	Joint Surveillance Target Attack Radar System
LCAC	landing craft air cushion
LRIP	low-rate initial production
OSD	Office of the Secretary of Defense
RSIP	radar system improvement program
USD (A&T)	Under Secretary of Defense (Acquisition and Technology)

Scope and Methodology

To develop information for this report, we selected a case study methodology—evaluating the conduct and practices of DOT&E through an analysis of 13 weapon systems. Recognizing that many test and evaluation issues are unique to individual systems, we determined that a case study methodology would offer the greatest probability of illuminating the variety of factors that impact the value or effectiveness of oversight at the level of the Office of the Secretary of Defense (OSD). Moreover, with nearly 200 systems subject to review of the Director, Operational Test and Evaluation (DOT&E) at any one time, we sought a sample that would enable us to determine if the Office had any impact as well as the ability to examine the variety of programs overseen. Therefore, we selected a judgmental sample of cases reflecting the breadth of program types. As illustrated in table I.1,¹ we selected systems (1) from each of the primary services, (2) categorized as major defense systems, and (3) representing a wide array of acquisition and testing phases—from early operational assessments through and beyond the full-rate production decision. We studied both new and modified systems.

Table I.1: Characteristics of Weapon Systems Used for Case Studies

System	Service(s)	Acquisition category ^a	Estimated or actual year of selected development phase	New or modification of existing system
AH-64D Longbow Apache helicopter	Army	1D	MS III (1995); IOT&E (1995)	Modification
Airborne Self-Protection Jammer	Navy	1D	FOT&E (1995-96); Bosnia (1995)	New
C-17A Airlifter	Air Force	1D	FOT&E (1996-98); MS IIIB (1995)	New
E-3 AWACS Radar System Improvement Program	Air Force	1C	AFSARC III (1997); IOT&E (1995-96)	Modification
F-22 fighter aircraft	Air Force	1D	MS III (2003); IOT&E (2002); LRIP (1999)	New
Javelin missile	Army	1D	MS III (1997); LUT (1996); UE (1996)	New

(continued)

¹Table I.1 lists the lead service, program size, and acquisition or testing phase for each of the case study systems, as well as whether the program is a development effort or a modification of an existing system.

Appendix I
Scope and Methodology

System	Service(s)	Acquisition category^a	Estimated or actual year of selected development phase	New or modification of existing system
Joint Surveillance Target Attack Radar System				
E-8 aircraft	Air Force	1D	FOT&E (1997); MS III (1996); Bosnia (1996)	New
Common ground station	Army	1D	MS III (1998); IOT&E (1997-98); Bosnia (1995)	New
LPD-17 Amphibious Assault Ship	Navy	1D	MS II (1996); EOA-2 (1996); EOA-1 (1994-95)	New
M1A2 tank	Army	1D	FOT&E (1995-96); MS III (1994)	Modification
Sensor Fuzed Weapon	Air Force	1D	FOT&E-1 (1997-98); MS III (1996); IOT&E-2 (1995-96)	New
Standard Missile SM-2				
Block IIIB version	Navy	II	FOT&E (1997); MS III (1996); OPEVAL (1996)	Modification
Block IV version	Navy	1D	MS III (1997); DT/IOT&E (1994)	Modification
Tomahawk Weapon System				
Baseline III	Navy	1C	MS III (1998); OPEVAL (1998); IOT&E (1997)	Modification
Baseline IV	Navy	1C	MS III (2000); OPEVAL (1999-00); IOT&E (1999)	Modification
V-22	Navy	1D	OPEVAL (1999); OT-IIC (1996)	New

(Table notes on next page)

Appendix I

Scope and Methodology

Legend

AFSARC = Air Force Systems Acquisition Review Council
DT = developmental testing
EOA = early operational assessment
FOT&E = follow-on operational test and evaluation
IOT&E = initial operational test and evaluation
LRIP = low-rate initial production
LUT = limited user test
MS = milestone
OA = operational assessment
OPEVAL = operational evaluation
OT = operational testing
UE = user evaluation

^aThe Under Secretary of Defense for Acquisition and Technology (USD (A&T)) designates major defense acquisition programs as either acquisition category 1D or 1C. The milestone decision authority for category 1D programs is USD (A&T). The milestone decision authority for category 1C programs is the Department of Defense (DOD) component head or, if delegated, the DOD component acquisition executive. Category I programs are major defense acquisition programs estimated to require more than \$355 million (fiscal year 1996 constant dollars) for expenditures in research, development, test, and evaluation, or more than \$2.135 billion (fiscal year 1996 constant dollars) for procurement. Category II programs are those that do not meet the criteria for category I but do meet the criteria for a major system. A major system is estimated to require more than \$75 million in fiscal year 1980 constant dollars (approximately \$140 million in fiscal year 1996 constant dollars) for expenditures in research, development, test, and evaluation, or more than \$300 million in fiscal year 1980 constant dollars (approximately \$645 million in fiscal year 1996 constant dollars) for procurement.

DOT&E, the service operational test agencies, and the Institute for Defense Analyses (IDA) personnel agreed that DOT&E was influential in the testing done on these 13 systems. In several cases, the participating agencies vehemently differed on the value of DOT&E's actions; however, whether DOT&E had an impact on testing (be it perceived as positive or negative) was not in dispute.

In conducting our 13 case studies, we assessed the strengths and weaknesses of the organizational framework in DOD for operational testing via test agency representatives, an assessment on the origins and implementation (exemplified by the 13 cases) of the title 10 amendments creating and empowering DOT&E, and a review of the literature.

To compile case study data, we interviewed current action officers in both DOT&E and the appropriate operational test agency and reviewed documentation provided by the operational test agencies, DOT&E, and IDA. Using structured questionnaires, we interviewed 12 DOT&E and 27 operational test agency action officers responsible for the 13 selected systems as well as managers and technical support personnel in each organization. In addition, we interviewed the commanders of each of the

service testing agencies and DOT&E. When possible, we corroborated information obtained from interviews with documentation, including test and evaluation master plans, beyond low-rate initial production reports, defense acquisition executive summary status reports, defense acquisition memoranda, and interagency correspondence.

In Washington, D.C., we obtained data from or performed work at the Office of the Director of Operational Test and Evaluation, OSD; Deputy Under Secretary of Defense for Acquisition Reform; Directorate of Navy Test and Evaluation and Technology Requirements, Office of the Chief of Naval Operations; Test and Evaluation Management Agency, Director of Army Staff; Air Force Test and Evaluation Directorate; and the DOD Office of the Inspector General. We also reviewed data and interviewed officials from the Army Operational Test and Evaluation Command and the Institute for Defense Analyses, Alexandria, Virginia; the Navy Commander, Operational Test and Evaluation Force, Norfolk, Virginia; and the Air Force Operational Test and Evaluation Command, Kirtland Air Force Base, New Mexico.

The use of a systematic case study framework enabled us to identify and categorize the types of impacts attributable to DOT&E among the systems studied. In addition, this framework enabled us to identify trends among factors that correlate with DOT&E effectiveness. However, we were unable to generalize to all systems subject to OSD-level oversight. In light of this limitation, we included only major (high-cost) systems and systems identified by DOT&E and the lead operational test agency as having been affected by DOT&E initiatives. Moreover, while our methodology and data collection enabled us to qualitatively assess the impact of DOT&E, it was not sufficiently rigorous either to evaluate the cost-effectiveness of DOT&E actions or to determine the deterrent effects, if any, the Office exerts over the acquisition and testing process. Finally, our methodology did not enable an assessment of whether the additional testing requested by DOT&E was necessary to provide full-rate production decisionmakers the essential information on a system's operational effectiveness and suitability or whether the additional data was worth the time, expense, and resources necessary to obtain it.

Our review was performed from June 1996 through March 1997 in accordance with generally accepted government auditing standards.

Description of 13 Case Study Systems

AH-64D Longbow Apache Helicopter

The AH-64D Longbow Apache is a remanufactured and upgraded version of the AH-64A Apache helicopter. This Army system is equipped with a mast-mounted fire control radar, fire-and-forget radio frequency Hellfire missile, and airframe improvements (i.e., integrated cockpit, improved engines, and global positioning system navigation).

Airborne Self-Protection Jammer

The Airborne Self-Protection Jammer is a defensive electronic countermeasures system using reprogrammable deceptive jamming techniques to protect tactical aircraft from radar-guided weapons. This Navy system is intended to protect Navy and Marine Corps F-18 and F-14 aircraft.

C-17A Airlifter

The C-17A Airlifter provides strategic/tactical transport of all cargo, including outsized cargo, mostly to main operational bases or to small, austere airfields, if needed. Its four-engine turbofan design enables the transport of large payloads over intercontinental ranges without refueling. This Air Force aircraft will replace the retiring C-141 aircraft and augment the C-130 and C-5 transport fleets.

E-3 AWACS Radar System Improvement Program

The Air Force's E-3 AWACS consists of a Boeing 707 airframe modified to carry a radome housing a pulse-Doppler radar capable of detecting aircraft and cruise missiles, particularly at low altitudes. The Radar System Improvement Program replaces several components of the radar to improve detection capability and electronic countermeasures as well as reliability, availability, and maintainability.

F-22 Air Superiority Fighter

The F-22 is an air superiority aircraft with a capability to deliver air-to-ground weapons. The most significant features include supercruise, the ability to fly efficiently at supersonic speeds without using fuel-consuming afterburners, low observability to adversary systems with the goal to locate and shoot down the F-22, and integrated avionics to significantly improve the pilot's battlefield awareness.

Javelin Missile

The Javelin is a man-portable, antiarmor weapon developed for the Army and the Marine Corp to replace the aging Dragon system. It is designed as a fire-and-forget system comprised of a missile and reusable command launch unit.

Joint Surveillance Target Attack Radar System

The Joint Surveillance Target Attack Radar System is designed to provide intelligence on moving and stationary targets to Air Force and Army command nodes in near real time. The system comprises a modified Boeing 707 aircraft frame equipped with radar, communications equipment, and the air component of the data link, computer workstations, and self-defense suite as well as ground station modules mounted on Army vehicles.

LPD-17 Amphibious Assault Ship

The LPD-17 will be an amphibious assault ship capable of launching (1) amphibious assault craft from a well deck and (2) helicopters or vertical takeoff and landing aircraft from an aft flight deck. It is intended to transport and deploy combat and support elements of Marine expeditionary brigades as a key component of amphibious task forces.

M1A2 Abrams Main Battle Tank

The M1A2 Abrams main battle tank is an upgrade of the M1A1 and is intended to improve target acquisition and engagement rates and survivability while sustaining equivalent operational suitability. Specifically, the modified tank incorporates a commander's independent thermal viewer, a position navigation system, and an intervehicle command and control system.

Sensor Fuzed Weapon

The Sensor Fuzed Weapon is an antiarmor cluster munition to be employed by fighter, attack, or bomber aircraft to achieve multiple kills per pass against armored and support combat formations. Each munition contains a tactical munitions dispenser comprising 10 submunitions containing a total of 40 infrared sensing projectiles. High-altitude accuracy is to be improved through the incorporation of a wind-compensated munition dispenser upgrade.

Standard Missile-2

The Standard Missile-2 is a solid propellant-fueled, tail-controlled, surface-to-air missile fired by surface ships. It was originally designed to counter high-speed, high-altitude antiship missiles in an advanced electronic countermeasures environment. The block IIIA version provides improved capacity against low-altitude targets with an improved warhead. The block IIIB adds an infrared seeker to the block IIIA to enhance the missile's capabilities against specific threats. These improvements are being made to provide capability against theater ballistic missiles while retaining its capabilities against antiair warfare threats.

Tomahawk Weapon System

The Tomahawk Weapon System is a long-range subsonic cruise missile for land and sea targets. The baseline IV upgrade is fitted with a terminal seeker, video data link, and two-way digital data link. The primary baseline IV configuration is the Tomahawk multimission missile; a second variant is the Tomahawk hard target penetrator.

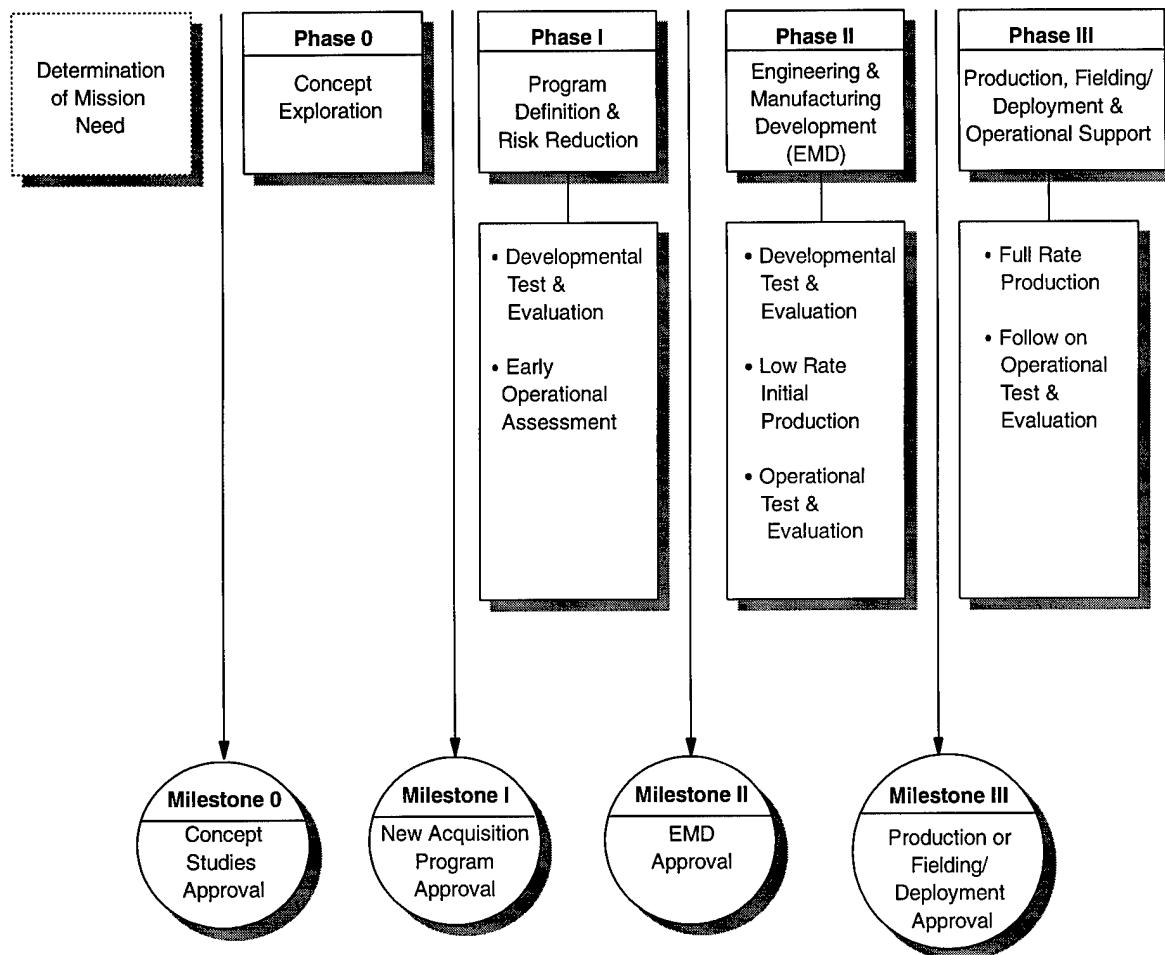
V-22 Osprey

The V-22 is a tilt rotor vertical/short takeoff and landing, multimission aircraft developed to fulfill operational combat requirements in the Marine Corps and Special Operations Forces.

DOT&E's and DOD's System Acquisition Process

DOT&E's role in the system acquisition process does not become prominent until the latter stages. As weapon system programs progress through successive phases of the acquisition process, they are subject to major decision points called milestones. The milestone review process is predicated on the principle that systems advance to higher acquisition phases by demonstrating that they meet prescribed technical and performance thresholds. Figure III.1 illustrates DOD's weapon system acquisition process.

Figure III.1: DOD's Weapon System Acquisition Process



Appendix III
**DOT&E's and DOD's System Acquisition
Process**

Per DOD directive, test and evaluation planning begins in phase 0, Concept Exploration. Operational testers are to be involved early to ensure that the test program for the most promising alternative can support the acquisition strategy and to ensure the harmonization of objectives, thresholds, and measures of effectiveness in the operational readiness document and the test and evaluation master plan. Early testing of prototypes in phase I, Program Definition and Risk Reduction, and early operational assessments are to be emphasized to assist in identifying risks. A combined developmental and operational test approach is encouraged to save time and costs. Initial operational test and evaluation is to occur during phase II to evaluate operational effectiveness and suitability before the full-rate production decision, milestone III, on all acquisition category I and II programs. For all acquisition category I programs and other programs designated for OSD test and evaluation oversight, a test and evaluation master plan is prepared and submitted for approval prior to first milestone review (excluding milestone 0).¹ The master plan is to be updated at milestones when the program has changed significantly. DOT&E must approve the test and evaluation master plan and the more specific operational test plans prior to their execution. This process and the required plan approvals provide DOT&E opportunities to affect the design and execution of operational testing throughout the acquisition process.

¹Master plans for acquisition category I programs are to be submitted to the Director, Test Systems Engineering and Evaluation, 30 days prior to the first milestone. For all other programs designated for OSD oversight, the plans must be submitted 90 days prior to the first milestone.

Comments From the Department of Defense

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

<p>See comment 1.</p> <p>See comment 2.</p> <p>See comment 3.</p>	<p>OFFICE OF THE SECRETARY OF DEFENSE WASHINGTON, DC 20301-1700</p> <p> OPERATIONAL TEST AND EVALUATION</p> <p>10 SEP 1997</p> <p>Mr. Kwai-Cheung Chan Director, Special Studies and Evaluation National Security and International Affairs Division U.S. General Accounting Office Washington, DC 20548</p> <p>Dear Mr Chan:</p> <p>This is the Department of Defense (DoD) response to the General Accounting Office (GAO) draft report, "TEST AND EVALUATION: Impact of DoD's Office of the Director of Operational Test and Evaluation," dated August 22, 1997, (GAO CODE 973444/OSD Case 1443). DoD partially concurs with the recommendations in the draft report.</p> <p>The DoD is committed to fielding weapon systems that substantially improve the warfighters' capabilities in a timely and affordable manner. In the current climate of improving efficiency, some departures from the past ways of doing business are warranted. The existence of the DOT&E organization, closely involved with, but independent of the acquisition community, helps ensure that we can improve the way we acquire effective, suitable, and survivable new systems. The information gained from well-run, independent OT&E activities is being heard and considered as part of the acquisition decision-making process.</p> <p>The discussion in the GAO Report on the Secretary of Defense initiatives for operational test and evaluation does not do justice to these important themes. These initiatives, especially the early involvement of operational testers in acquisition programs, are producing earlier insights on the performance of military systems. It makes no sense to wait until Milestone III to discover problems which could have been learned and corrected years earlier.</p> <p>The GAO also discusses the small DOT&E test resource management planning and leadership role in the context that it may detract from testing oversight missions. Since the financial and human test resources budgeted by the Services have been declining, the very ability of the Services to perform adequate operational testing is at stake. As GAO notes, DOT&E is responsible by statute to review and make recommendations to the Secretary of Defense on all budgetary matters relating to operational testing. Therefore it is surprising that GAO considers DOT&E's efforts in test resources as potentially impacting the adequacy of oversight rather than contributing to the quality of operational test and evaluation in DoD.</p>
-------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Appendix IV
Comments From the Department of Defense

Detailed responses to each of the GAO recommendations are contained in the attachment
DoD appreciates the opportunity to review and comment on the draft report.



Philip E. Coyle
Director

Attachment:
As stated

**GAO DRAFT REPORT DATED AUGUST 22, 1997
(GAO CODE 973444) OSD CASE 1443**

**"TEST AND EVALUATION: IMPACT OF DOD'S OFFICE OF THE DIRECTOR OF
OPERATIONAL TEST AND EVALUATION"**

**DEPARTMENT OF DEFENSE COMMENTS ON
THE GAO RECOMMENDATIONS**

RECOMMENDATION 1: The GAO recommended that the Secretary of Defense require the Under Secretary of Defense for Acquisition and Technology, in those cases where affirmative full-rate production decisions are made for systems that have yet to demonstrate their operational effectiveness or suitability, to (1) take corrective actions to eliminate effectiveness and suitability deficiencies, and (2) conduct follow-on test and evaluation of corrective actions until the systems are determined to be operationally effective and suitable by the Director, Operational Test and Evaluation. (p. 32/GAO Draft Report)

DOD RESPONSE: Concur. In cases where affirmative full-rate production decisions are made for systems that have yet to demonstrate their operational effectiveness or suitability, the milestone decision authority may, under current practice, proceed with production and fielding if other conditions warrant. Commensurate with schedule and resource constraints, serious deficiencies remaining at the time of the decision to proceed beyond low-rate initial production assume a high priority for correction and re-testing to ensure that the user's requirements are met.

By statute and DoD Directive, the Director, Operational Test and Evaluation (DOT&E), is the senior advisor to the Secretary of Defense and the Defense Acquisition Executive (DAE) on operational test and evaluation matters. Before the Milestone III decision, the program Test and Evaluation Master Plan (TEMP) must be approved by DOT&E. At that time, the plans for FOT&E must be adequate to ensure that corrections to previous deficiencies are thoroughly tested and evaluated.

RECOMMENDATION 2: The GAO recommended that the Secretary of Defense require the Director, Operational Test and Evaluation, to (1) review and approve follow-on test and evaluation master plans and specific operational test plans prior to the conduct of operational testing related to suitability and effectiveness issues left unresolved at the full-rate production decision, and (2) report to the Secretary of Defense, the Under Secretary of Defense for Acquisition and Technology, and the Congress upon the completion of follow-on operational test and evaluation whether the testing was adequate and whether the results confirmed the system is operationally suitable and effective. (p. 33/GAO Draft Report)

Now on p. 22.

See comment 4.

Now on p. 22.

Appendix IV
Comments From the Department of Defense

See comment 5.

Now on p. 23.

DOD RESPONSE: Partially concur. Title 10, USC, Section 139, states that the Director shall monitor and review all OT&E within the DoD. As a practical matter, DOT&E requires TEMP/test plan approval, test monitoring, and results reporting for FOT&E, as described in the Director, OT&E, policy memorandum of March 10, 1997, for programs subject to oversight. DOT&E's normal practice is to include such test results in the next DOT&E Annual Report to Congress. DOT&E already submits an independent report to the Secretary of Defense and the congressional committees following initial OT&E, and before the decision to proceed beyond low-rate initial production is finalized. DoD does not believe that a system-specific FOT&E report to the Secretary and Congress is warranted under most circumstances, given the resource limitations present, and lacking extraordinary requirements for such a report.

RECOMMENDATION 3: The GAO recommended, in light of increasing operational testing oversight commitments and to accommodate oversight of follow-on operational testing and evaluation (FOT&E) that DOT&E prioritize the office's workload to ensure sufficient attention is given to major defense acquisition programs. (p. 33/GAO Draft Report)

DOD RESPONSE: Concur. Acquisition reform has caused a great deal of change throughout the range of acquisition-related activities, including test and evaluation. Much of this change has increased the workload of DOT&E staff, but our first priority will continue to be to fulfill statutory requirements, including the oversight of major defense acquisition programs' OT&E and live fire T&E. The prioritization of OT&E activities is an ongoing and continuous process by the Director, his Deputies, and their Action Officers. The participation by DOT&E Action Officers in integrated product/process team activities prevalent in acquisition programs, while demanding of time, results in efficient communication among diverse organizations involved with test planning, conduct, and evaluation. DoD believes that through this improved communication, better-integrated and more efficient test programs will result at lower cost, and less time. Recognizing the demands of acquisition reform and the results of the Quadrennial Defense Review, DOT&E has requested resources to match these responsibilities.

The following are GAO's comments on the September 19, 1997, letter from the Department of Defense.

GAO Comments

1. In prior reviews of individual weapon systems, we have found that operational testing and evaluation is generally viewed by the acquisition community as a costly and time-consuming requirement imposed by outsiders rather than a management tool for more successful programs. Efforts to enhance the efficiency of acquisition, in general—and in operational testing, in particular—need to be well balanced with the requirement to realistically and thoroughly test operational suitability and effectiveness prior to the full-rate production decision. We attempted to take a broader view of acquisition reform efficiency initiatives to anticipate how these departures from past ways of doing business could impact both the quality of operational testing and the independence of DOT&E.
2. We were asked to assess the impact of DOT&E on the quality and impact of testing and reported on the Secretary of Defense initiatives only to the extent they may pose a potential impact on DOT&E's independence or effectiveness. Moreover, we did not recommend or suggest that testers wait until milestone III to discover problems that could have been learned and corrected earlier. Since its inception, DOT&E has been active in test integration and planning working groups and test and evaluation master plan development during the earliest phases of the acquisition process. In fact, we have long advocated more early testing to demonstrate positive system performance prior to the low-rate initial production decision. DOT&E's early involvement in test planning is appropriate, necessary, and required by DOD regulations. In this report we do not advocate the elimination of DOT&E participation during the early stages of the acquisition process; rather, we merely observe that DOT&E participation through the vehicle of working-level program manager integrated product teams has the potential to complicate independence and may be increasingly difficult to implement with declining resources and increasing oversight responsibilities following milestone III.
3. We did not recommend or suggest that DOT&E ignore its statutory responsibility to review and make recommendations to the Secretary of Defense on budgetary and financial matters related to operational test facilities and equipment. We only observed that in an era of declining resources, earlier participation, and extended oversight responsibilities, a decision to assume a larger role in test resource management planning and

leadership is likely to result in tradeoffs in other responsibilities—the largest being oversight.

4. We made this recommendation because DOT&E, the services, and the program offices did not necessarily agree on the degree to which system performance requirements have been met in initial operational test and evaluation. Furthermore, there was no consensus within the acquisition community concerning DOT&E's authority to oversee follow-on operational test and evaluation conducted to ensure that proposed corrections to previously identified deficiencies were thoroughly tested and evaluated.
5. Under 10 U.S.C. 2399, DOT&E is required to independently report to Congress whether a major acquisition system has proven to be operationally suitable and effective prior to the full-rate production decision. When follow-on operational test and evaluation is necessary to test measures intended to correct deficiencies identified in initial operational test and evaluation, Congress does not receive an equivalent independent report from DOT&E that concludes, based on required follow-on operational test and evaluation, whether or not a major system has improved sufficiently to be considered both operationally suitable and effective.

Appendix IV
Comments From the Department of Defense

Related GAO Products

Tactical Intelligence: Joint STARS Full-Rate Production Decision Was Premature and Risky (GAO/NSIAD-97-68, Apr. 25, 1997).

Weapons Acquisition: Better Use of Limited DOD Acquisition Funding Would Reduce Costs (GAO/NSIAD-97-23, Feb. 13, 1997).

Airborne Self-Protection Jammer (GAO/NSIAD-97-46R, Jan. 29, 1997).

Army Acquisition: Javelin Is Not Ready for Multiyear Procurement (GAO/NSIAD-96-199, Sept. 26, 1996).

Tactical Intelligence: Accelerated Joint STARS Ground Station Acquisition Strategy Is Risky (GAO/NSIAD-96-71, May 23, 1996).

Electronic Warfare (GAO/NSIAD-96-109R, Mar. 1, 1996).

Longbow Apache Helicopter: System Procurement Issues Need to Be Resolved (GAO/NSIAD-95-159, Aug. 24, 1995).

Electronic Warfare: Most Air Force ALQ-135 Jammers Procured Without Operational Testing (GAO/NSIAD-95-47, Nov. 22, 1994).

Weapons Acquisition: Low-Rate Initial Production Used to Buy Weapon Systems Prematurely (GAO/NSIAD-95-18, Nov. 21, 1994).

Acquisition Reform: Role of Test and Evaluation in System Acquisition Should Not Be Weakened (GAO/T-NSIAD-94-124, Mar. 22, 1994).

Test and Evaluation: The Director, Operational Test and Evaluation's Role in Test Resources (GAO/NSIAD-90-128, Aug. 27, 1990).

Adequacy of Department of Defense Operational Test and Evaluation (GAO/T-NSIAD-89-39, June 16, 1989).

Weapons Testing: Quality of DOD Operational Testing and Reporting (GAO/PEMD-88-32BR, July 26, 1988).

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are \$2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

**U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013**

or visit:

**Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC**

**Orders may also be placed by calling (202) 512-6000
or by using fax number (202) 512-6061, or TDD (202) 512-2537.**

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

**For information on how to access GAO reports on the INTERNET,
send an e-mail message with "info" in the body to:**

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

<http://www.gao.gov>